(72) Inventors: WOLD, Erling, H.; 5618 Ludwig Avenue, El Cerrito, CA 94530 (US). BLUM, Thomas, L.; 1740 Franklin Street, No. 2, San Francisco, CA 94109 (US). KEISLAR, Douglas, F.; 779 Hilldate Avenue, Berkeley, CA 94708 (US). WHEATON, James, A.; 30 Scenic Road, Fairfax, CA 94930 (US).

(74) Agents: D'ALESSANDRO, Kenneth et al.; Sierra Patent Group, Ltd., P.O. Box 6149, Stateline, NV 89449 (US).

(54) Title: METHOD AND APPARATUS FOR CREATING A UNIQUE AUDIO SIGNATURE

(57) Abstract: A method and apparatus for creating a signature of a sampled work in real-time is disclosed herein. Unique signatures of an unknown audio work are created by segmenting a file into segments having predetermined segment and hop sizes. The signature then may be compared against reference signatures. One aspect may be characterized in that the hop size of the sampled work signature is less than the hop size of reference signatures. A method for identifying an unknown audio work is also disclosed.

SPECIFICATION
METHOD AND APPARATUS FOR
CREATING A UNIQUE AUDIO SIGNATURE

10   BACKGROUND OF THE INVENTION

Field of the Invention

The present invention relates to data communications. In particular, the present
invention relates to creating a unique audio signature.

The Prior Art

15   Background

Digital audio technology has greatly changed the landscape of music and
entertainment. Rapid increases in computing power coupled with decreases in cost have
made it possible individuals to generate finished products having a quality once available
only in a major studio. Once consequence of modern technology is that legacy media

20   storage standards, such as reel-to-reel tapes, are being rapidly replaced by digital storage
media, such as the Digital Versatile Disk (DVD), and Digital Audio Tape (DAT).
Additionally, with higher capacity hard drives standard on most personal computers, home
users may now store digital files such as audio or video tracks on their home computers.

Furthermore, the Internet has generated much excitement, particularly among those

25   who see the Internet as an opportunity to develop new avenues for artistic expression and
communication. The Internet has become a virtual gallery, where artists may post their
works on a Web page. Once posted, the works may be viewed by anyone having access to
the Internet.

One application of the Internet that has received considerable attention is the ability

30   to transmit recorded music over the Internet. Once music has been digitally encoded into a
file, the file may be both downloaded by users for play, or broadcast ("streamed") over the
Internet. When files are streamed, they may be listened to by Internet users in a manner
much like traditional radio stations.

Given the widespread use of digital media, digital audio files, or digital video files

35   containing audio information, may need to be identified. The need for identification of
digital files may arise in a variety of situations. For example, an artist may wish to verify
royalty payments or generate their own Arbitron®-like ratings by identifying how often their
works are being streamed or downloaded. Additionally, users may wish to identify a a

1

5

particular work. The prior art has made efforts to create methods for identifying digital
audio works.

However, systems of the prior art suffer from certain disadvantages. For example,
prior art systems typically create a reference signature by examining the copyrighted work as
10   a whole, and then creating a signature based upon the audio characteristics of the entire
work. However, examining a work in total can result in a signature may not accurately
represent the original work. Often, a work may have distinctive passages which may not be
reflected in a signature based upon the total work. Furthermore, often works are
electronically processed prior to being streamed or downloaded, in a manner that may affect
15   details of the work's audio characteristics, which may result in prior art systems missing the
identification of such works. Examples of such electronic processing include data
compression and various sorts of audio signal processing such as equalization.

Hence, there exists a need to provide a system which overcomes the disadvantages of
the prior art.

20   BRIEF DESCRIPTION OF THE INVENTION

The present invention relates to data communications. In particular, the present
invention relates to creating a unique audio signature.

A method for creating a signature of a sampled work in real-time is disclosed herein.
One aspect of the present invention comprises: receiving a sampled work; segmenting the
25   sampled work into a plurality of segments, the segments having predetermined segment and
hop sizes; creating a signature of the sampled work based upon the plurality of segments;
and storing the sampled work signature. Additional aspects include providing a plurality of
reference signatures having a segment size and a hop size. An additional aspect may be
characterized in that the hop size of the sampled work signature is less than the hop size of
30   the reference signatures.

An apparatus for creating a signature of a sampled work in real-time is also
disclosed. In a preferred aspect, the apparatus comprises: means for receiving a sampled
work; means for segmenting the sampled work into a plurality of segments, the segments
having predetermined segment and hop sizes; means for creating a signature of the sampled
35   work based upon the plurality of segments; and storing the sampled work signature.
Additional aspects include means for providing a plurality of reference signatures having a

5　　segment size and a hop size. An additional aspect may be characterized in that the hop size

of the sampled work signature is less than the hop size of the reference signatures.

A method for identifying an unknown audio work is also disclosed. In another aspect

of the present invention, the method comprises: providing a plurality of reference signatures

10　　each having a segment size and a hop size; receiving a sampled work; creating a signature of

the sampled work, the sampled work signature having a segment size and a hop size; storing

the sampled work signature; comparing the sampled work signature to the plurality of

reference signatures to determine whether there is a match; and wherein the method is

characterized in that the hop size of the sampled work signature is less than the hop size of

15　　the reference signatures.

Further aspects of the present invention include creating a signature of the sampled

work by calculating segment feature vectors for each segment of the sampled work. The

segment feature vectors may include MFCCs calculated for each segment.

BRIEF DESCRIPTION OF THE DRAWING FIGURES

20　　Figure 1 is a flowchart of a method according to the present invention.

Figure 2 is a diagram of a system suitable for use with the present invention.

Figure 3 is a diagram of segmenting according to the present invention.

Figure 4 is a detailed diagram of segmenting according to the present invention

showing hop size.

25　　Figure 5 is a graphical flowchart showing the creating of a segment feature vector

according to the present invention.

Figure 6 is a diagram of a signature according to the present invention.

Figure 7 is a functional diagram of a comparison process according to the present

invention.

30　　DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

Persons of ordinary skill in the art will realize that the following description of the

present invention is illustrative only and not in any way limiting. Other embodiments of the

invention will readily suggest themselves to such skilled persons having the benefit of this

disclosure.

35　　It is contemplated that the present invention may be embodied in various computer

and machine-readable data structures. Furthermore, it is contemplated that data structures

embodying the present invention will be transmitted across computer and machine-readable

5   media, and through communications systems by use of standard protocols such as those used

to enable the Internet and other computer networking standards.

The invention further relates to machine-readable media on which are stored

embodiments of the present invention. It is contemplated that any media suitable for storing

10   instructions related to the present invention is within the scope of the present invention. By

way of example, such media may take the form of magnetic, optical, or semiconductor

media.

The present invention may be described through the use of flowcharts. Often, a

single instance of an embodiment of the present invention will be shown. As is appreciated

15   by those of ordinary skill in the art, however, the protocols, processes, and procedures

described herein may be repeated continuously or as often as necessary to satisfy the needs

described herein. Accordingly, the representation of the present invention through the use of

flowcharts should not be used to limit the scope of the present invention.

The present invention may also be described through the use of web pages in which

20   embodiments of the present invention may be viewed and manipulated. It is contemplated

that such web pages may be programmed with web page creation programs using languages

standard in the art such as HTML or XML. It is also contemplated that the web pages

described herein may be viewed and manipulated with web browsers running on operating

systems standard in the art, such as the Microsoft Windows® and Macintosh® versions of

25   Internet Explorer® and Netscape®. Furthermore, it is contemplated that the functions

performed by the various web pages described herein may be implemented through the use

of standard programming languages such a Java® or similar languages.

The present invention will first be described in general overview. Then, each

element will be described in further detail below.

30   Referring now to Figure 1, a flowchart is shown which provides a general overview

of the present invention. The present invention may be viewed as three steps: 1) receiving a

sampled work; 2) segmenting the work; 3) creating signatures of the segments; and 4)

storing the signatures of the segments.

Receiving a sampled work

35   Beginning with act 100, a sampled work is provided to the present invention. It is

contemplated that the work will be provided to the present invention as a digital audio

stream.

5

It should be understood that if the audio is in analog form, it may be digitized in a manner standard in the art.

Segmenting the work

After the sampled worked is received, the work is then segmented in act 102. It is
10   contemplated that the sampled work may be segmented into predetermined lengths. Though segments may be of any length, the segments of the present invention are preferably of the same length.

In an exemplary non-limiting embodiment of the present invention, the segment lengths are in the range of 0.5 to 3 seconds. It is contemplated that if one were searching for
15   very short sounds (e.g., sound effects such as gunshots), segments as small as 0.01 seconds may be used in the present invention. Since humans don't resolve audio changes below about 0.018 seconds, segment lengths less than 0.018 seconds may not be useful. On the other hand, segment lengths as high as 30-60 seconds may be used in the present invention. The inventors have found that beyond 30-60 seconds may not be useful, since most details in
20   the signal tend to average out.

Generating signatures

Next, in act 104, each segment is analyzed to produce a signature, known herein as a segment feature vector. It is contemplated that a wide variety of methods known in the art may be used to analyze the segments and generate segment feature vectors. In an exemplary
25   non-limiting embodiment of the present invention, the segment feature vectors may be created using the method described in US Patent #5,918,223 to Blum, et al, which is incorporated by reference as though set forth fully herein.

Storing the signatures

In act 106, the segment feature vectors are stored to create a representative signature
30   of the sampled work.

Each above-listed step will now be shown and described in detail.

Referring now to Figure 2, a diagram of a system suitable for use with the present invention is shown. FIG. 2 includes a client system 200. It is contemplated that client system 200 may comprise a personal computer 202 including hardware and software
35   standard in the art to run an operating system such as Microsoft Windows ®, MAC OS ®, or other operating systems standard in the art. Client system 200 may further include a database 204 for storing and retrieving embodiments of the present invention. It is

5

contemplated that database 204 may comprise hardware and software standard in the art and may be operatively coupled to PC 202. Database 204 may also be used to store and retrieve the works and segments utilized by the present invention.

Client system 200 may further include an audio/video (A/V) input device 208. A/V

10   device 208 is operatively coupled to PC 202 and is configured to provide works to the present invention which may be stored in traditional audio or video formats. It is contemplated that A/V device 208 may comprise hardware and software standard in the art configured to receive and sample audio works (including video containing audio information), and provide the sampled works to the present invention as digital audio files.

15   Typically, the A/V input device 208 would supply raw audio samples in a format such as 16-bit stereo PCM format. A/V input device 208 provides an example of means for receiving a sampled work.

It is contemplated that sampled works may be obtained over the Internet, also. Typically, streaming media over the Internet is provided by a provider, such as provider 218

20   of FIG. 2. Provider 218 includes a streaming application server 220, configured to retrieve works from database 222 and stream the works in a formats standard in the art, such as Real®, Windows Media®, or QuickTime®. The server then provides the streamed works to a web server 224, which then provides the streamed work to the Internet 214 through a gateway 216. Internet 214 may be any packet-based network standard in the art, such as IP,

25   Frame Relay, or ATM.

To reach the provider 218, the present invention may utilize a cable or DSL head end 212 standard in the art operatively, which is coupled to a cable modem or DSL modem 210 which is in turn coupled to the system's network 206. The network 206 may be any network standard in the art, such as a LAN provided by a PC 202 configured to run software standard

30   in the art.

It is contemplated that the sampled work received by system 200 may contain audio information from a variety of sources known in the art, including, without limitation, radio, the audio portion of a television broadcast, Internet radio, the audio portion of an Internet video program or channel, streaming audio from a network audio server, audio delivered to

35   personal digital assistants over cellular or wireless communication systems, or cable and satellite broadcasts.

5

        Additionally, it is contemplated that the present invention may be configured to

receive and compare segments coming from a variety of sources either stored or in real-time.

For example, it is contemplated that the present invention may compare a real-time

streaming work coming from streaming server 218 or A/V device 208 with a reference

10   segment stored in database 204.

        Figure 3 shows a diagram showing the segmenting of a work according to the present

invention. FIG. 3 includes audio information 300 displayed along a time axis 302. FIG. 3

further includes a plurality of segments 304, 306, and 308 taken of audio information 300

over some segment size T.

15   In an exemplary non-limiting embodiment of the present invention, instantaneous

values of a variety of acoustic features are computed at a low level, preferably about 100

times a second. Additionally, 10 MFCCs (cepstral coefficients) are computed for each

segment. It is contemplated that any number of MFCCs may be computed. Preferably, 5-

20 MFCCs are computed, however, as many as 30 MFCCs may be computed, depending on

20   the need for accuracy versus speed.

        In an exemplary non-limiting embodiment of the present invention, the segment-

level acoustical features comprise statistical measures as disclosed in the '223 patent of these

low-level features calculated over the length of each segment. The data structure may store

other bookkeeping information as well (segment size, hop size, item ID, UPC, etc).

25   As can be seen by inspection of FIG. 3, the segments 304, 306, and 308 may overlap

in time. This amount of overlap may be represented by measuring the time between the

center point of adjacent segments. This amount of time is referred to herein as the hop size

of the segments, and is so designated in FIG. 3. By way of example, if the segment length T

of a given segment is one second, and adjacent segments overlap by 50%, the hop size

30   would be 0.5 second.

        The hop size may be set during the development of the software. Additionally, the

hop sizes of the reference database and the real-time segments may be predetermined to

facilitate compatibility. For example, the reference signatures in the reference database may

be precomputed with a fixed hop and segment size, and thus the client applications should

35   conform to this segment size and have a hop size which integrally divides the reference

signature hop size. It is contemplated that one may experiment with a variety of segment

5   sizes in order to balance the tradeoff of accuracy with speed of computation for a given
       application.

              The inventors have found that by carefully choosing the hop size of the segments, the
       accuracy of the identification process may be significantly increased. Additionally, the
10   inventors have found that the accuracy of the identification process may be increased if the
       hop size of reference segments and the hop size of segments obtained in real-time are each
       chosen independently. The importance of the hop size of segments may be illustrated by
       examining the process for segmenting pre-recorded works and real-time works separately.

              Reference signatures
15   Prior to attempting to identify a given work, a reference database of signatures must
       be created. When building a reference database, a segment length having a period of less
       than three seconds is preferred. In an exemplary non-limiting embodiment of the present
       invention, the segment lengths have a period ranging from 0.5 seconds to 3 seconds. For a
       reference database, the inventors have found that a hop size of approximately 50% to 100%
20   of the segment size is preferred.

              It is contemplated that the reference signatures may be stored on a database such as
       database 204 as described above. Database 204 and the discussion herein provide an
       example of means for providing a plurality of reference signatures each having a segment
       size and a hop size.

25   Real-time signatures
              The choice of the hop size is important for real-time segments.

              Figure 4 shows a detailed diagram of a real-time segment according to the present
       invention. FIG. 4 includes real-time audio information 400 displayed along a time axis 402.
       FIG. 4 further includes segments 404 and 406 taken of audio information 400 over some
30   segment length T. In an exemplary non-limiting embodiment of the present invention, the
       segment length of real-time segments is chosen to range from 0.5 to 3 seconds.

              As can be seen by inspection of FIG. 4, the hop size of real-time is chosen to be
       smaller than that of reference segments. In an exemplary non-limiting embodiment of the
       present invention, the hop size of real-time segments is less than 50% of the segment size.
35   In yet another exemplary non-limiting embodiment of the present invention, the real-time
       hop size may be 0.1 seconds.

5   The inventors have found such a small hop size advantageous for the following
reasons. The ultimate purpose of generating real-time segments is to analyze and compare
them with the reference segments in the database to look for matches. The inventors have
found at least two major reasons why a segment of the same audio recording captured real-
10   time would not match its counterpart in the database. One is that the broadcast channel does
not produce a perfect copy of the original. For example, the work may be edited or
processed or the announcer may talk over part of the work. The other reason is that larger
segment boundaries may not line up in time with the original segment boundaries of the
target recordings.

15   The inventors have found that by choosing a smaller hop size, some of the segments
will ultimately have time boundaries that line up with the original segments, notwithstanding
the problems listed above. The segments that line up with a "clean" segment of the work
may then be used to make an accurate comparison while those that do not so line up may be
ignored. The inventors have found that a hop size of 0.1 seconds seems to be the maximum
20   that would solve this time shifting problem.

           As mentioned above, once a work has been segmented, the individual segments are
then analyzed to produce a segment feature vector. Figure 5 is a diagram showing an
overview of how the segment feature vectors may be created using the methods described in
US Patent #5,918,223 to Blum, et al. It is contemplated that a variety of analysis methods
25   may be useful in the present invention, and many different features may be used to make up
the feature vector. The inventors have found that the pitch, brightness, bandwidth, and
loudness features of the '223 patent to be useful in the present invention. Additionally,
spectral features may be used analyzed, such as the energy in various spectral bands. The
inventors have found that the cepstral features (MFCCs) are very robust (more invariant)
30   given the distortions typically introduced during broadcast, such as EQ, multi-band
compression/limiting, and audio data compression techniques such as MP3
encoding/decoding, etc.

           In act 500, the audio segment is sampled to produce a segment. In act 502, the
sampled segment is then analyzed using Fourier Transform techniques to transform the
35   signal into the frequency domain. In act 504, mel frequency filters are applied to the
transformed signal to extract the significant audible characteristics of the spectrum. In act
506, a Discrete Cosine Transform is applied which converts the signal into mel frequency

5   cepstral coefficients (MFCCs). Finally, in act 508, the MFCCs are then averaged over a predetermined period. In an exemplary non-limiting embodiment of the present invention, this period is approximately one second. Additionally, other characteristics may be computed at this time, such as brightness or loudness. A segment feature vector is then

10   produced which contains a list containing at least the 10 MFCCs corresponding average.

The disclosure of FIGS. 3, 4, and 5 provide examples of means for creating a signature of a sampled work having a segment size and a hop size.

Figure 6 is a diagram showing a complete signature 600 according to the present invention. Signature 600 includes a plurality of segment feature vectors 1 through n

15   generated as shown and described above. Signature 600 may also include an identification portion containing a unique ID. It is contemplated that the identification portion may contain a unique identifier provided by the RIAA (Recording Industry Association of America). The identification portion may also contain information such as the UPC (Universal Product Code) of the various products that contain the audio corresponding to

20   this signature. Additionally, it is contemplated that the signature 600 may also contain information pertaining to the characteristics of the file itself, such as the hop size, segment size, number of segments, etc., which may be useful for storing and indexing.

Signature 600 may then be stored in a database and used for comparisons. The following computer code in the C programming language provides an example of a

25   database structure in memory according to the present invention:

```
typedef struct
{
        float   hopSize;              /* hop size */
        float   segmentSize;          /* segment size */
30      MFSignature* signatures;      /* array of signatures */
} MFDatabase;
```

The following provides an example of the structure of a segment according to the present invention:

```
typedef struct
35      {
        char*   id;                   /* unique ID for this audio clip */
        long    numSegments;/* number of segments */
```

5

```
      float*  features;           /* feature array */
      long    size;               /* size of per-segment feature vector */
      float   hopSize;
      float   segmentSize;
```

10   } MFSignature;

The discussion of FIG. 6 provides an example of means for storing segments and signatures according to the present invention.

Figure 7 shows a functional diagram of a comparison process according to the present invention. Act 1 of FIG. 7 shows unknown audio being converted to a signature according to the present invention. In act 2, reference signatures are retrieved from a

15   reference database. Finally, the reference signatures are scanned and compared to the unknown audio signatures to determine whether a match exists. This comparison may be accomplished through means known in the art. For example, the Euclidean distance between the reference and real-time signature can be computed and compared to a threshold.

It is contemplated that the present invention has many beneficial uses, including

20   many outside of the music piracy area. For example, the present invention may be used to verify royalty payments. The verification may take place at the source or the listener. Also, the present invention may be utilized for the auditing of advertisements, or collecting Arbitron®-like data (who is listening to what). The present invention may also be used to

25   label the audio recordings on a user's hard disk or on the web.

While embodiments and applications of this invention have been shown and described, it would be apparent to those skilled in the art that many more modifications than mentioned above are possible without departing from the inventive concepts herein. The invention, therefore, is not to be restricted except in the spirit of the appended claims.

5

What is claimed is:

1.   A method for creating a signature of a sampled work in real-time, said method comprising:

        receiving a sampled work;

10   segmenting said sampled work into a plurality of segments, said segments having predetermined segment and hop sizes;

        creating a signature of said sampled work based upon said plurality of segments; and

        storing said sampled work signature.

15   2.   The method of claim 1, wherein said method further includes the act of providing a plurality of reference signatures having a segment size and a hop size.

3.   The method of claim 2, wherein said method is characterized in that said hop size of said sampled work signature is less than said hop size of said reference signatures.

4.   The method of claim 1, wherein said act of creating a signature of said sampled work

20   comprises calculating segment feature vectors for each segment of said sampled work.

5.   The method of claim 2, wherein said act of creating a signature includes calculating a plurality of MFCCs for each said segment.

6.   The method of claim 1, wherein said act of creating a signature includes calculating a plurality of acoustical features from the group consisting of at least one of loudness, pitch,

25   brightness, bandwidth, spectrum and MFCC coefficients for each said segment.

7.   The method of claim 1, wherein said sampled work signature comprises a plurality of segments and an identification portion.

8.   The method of claim 7, wherein said plurality of segments of said sampled work signature comprise a segment size of approximately 0.5 to 3 seconds.

30   9.   The method of claim 8, wherein said plurality of segments of said sampled work signature comprise a hop size of less than 50% of the segment size.

10.   The method of claim 8, wherein said plurality of segments of said sampled work signature comprise a hop size of approximately 0.1 seconds.

11.   A method for identifying an unknown audio work, said method comprising:

35   providing a plurality of reference signatures each having a segment size and a hop size;

        receiving a sampled work;

12

5   segmenting said sampled work into a plurality of segments, said segments

having predetermined segment and hop sizes;

creating a signature of said sampled work based upon said plurality of

segments;

10   storing said sampled work signature; and

comparing said sampled work signature to said plurality of reference

signatures to determine whether there is a match.

12.   The method of claim 11, wherein said method is characterized in that said hop size of

said sampled work signature is less than said hop size of said reference signatures.

15   13.   The method of claim 11, wherein said act of creating a signature of said sampled

work comprises calculating segment feature vectors for each segment of said sampled work.

14.   The method of claim 13, wherein said act of comparing said sampled work signature

to said plurality of reference signatures comprises measuring the Euclidean distance between

said feature vectors and comparing said Euclidean distance to a predetermined threshold.

20   15.   The method of claim 13, wherein said act of creating a signature includes calculating

a plurality of MFCCs for each said segment.

16.   The method of claim 13, wherein said act of creating a signature includes calculating

a plurality of acoustical features from the group consisting of at least one of loudness, pitch,

brightness, bandwidth, spectrum and MFCC coefficients for each said segment.

25   17.   The method of claim 11, wherein said sampled work signature comprises a plurality

of segments and an identification portion.

18.   The method of claim 17, wherein said plurality of segments of said sampled work

signature comprise a segment size of approximately 0.5 to 3 seconds.

19.   The method of claim 17, wherein said plurality of segments of said sampled work

30   signature comprise a hop size of less than 50% of the segment size.

20.   The method of claim 17, wherein said plurality of segments of said sampled work

signature comprise a hop size of approximately 0.1 seconds.

21.   An apparatus for creating a signature of a sampled work in real-time, said apparatus

comprising:

35   means for receiving a sampled work;

means for segmenting said sampled work into a plurality of segments, said

segments having predetermined segment and hop sizes;

13

means for creating a signature of said sampled work based upon said plurality of segments; and

means for storing said sampled work signature.

22.   The apparatus of claim 21, wherein said apparatus further includes means for providing a plurality of reference signatures having a segment size and a hop size.

23.   The apparatus of claim 22, wherein said apparatus is characterized in that said hop size of said sampled work signature is less than said hop size of said reference signatures.

24.   The apparatus of claim 21, wherein said means for creating a signature of said sampled work comprises calculating segment feature vectors for each segment of said sampled work.

25.   The apparatus of claim 21, wherein said means for creating a signature includes calculating a plurality of MFCCs for each said segment.

26   The apparatus of claim 21, wherein said sampled work signature comprises a plurality of segments and an identification portion.

27.   The apparatus of claim 26, wherein said plurality of segments of said sampled work signature comprise a segment size of approximately 0.5 to 3 seconds.

28.   The apparatus of claim 26, wherein said plurality of segments of said sampled work signature comprise a hop size of less than 50% of the segment size.

29.   The apparatus of claim 26, wherein said plurality of segments of said sampled work signature comprise a hop size of approximately 0.1 seconds.

30.   The apparatus of claim 21, wherein said means for creating a signature includes means for calculating a plurality of acoustical features from the group consisting of at least one of loudness, pitch, brightness, bandwidth, spectrum and MFCC coefficients for each said segment.

31.   A machine readable data transmission containing a data structure for a sampled work having a plurality of segments, said segments having predetermined segment and hop sizes.

32.   The machine readable data transmission of claim 31, wherein said data structure is characterized in that said hop size of said sampled work signature is less than said hop size of a corresponding reference signature.

33.   A machine-readable data medium containing a data structure for a sampled work having a plurality of segments, said segments having predetermined segment and hop sizes.

34.   The machine-readable data medium of claim 33, wherein said data structure is characterized in that said hop size of said sampled work signature is less than said hop size of a corresponding reference signature.

35.   A program storage device readable by a machine, tangibly embodying a program of instructions executable by the machine to perform a method for creating a signature of a sampled work in real-time, said method comprising:

  receiving a sampled work;

  segmenting said sampled work into a plurality of segments, said segments having predetermined segment and hop sizes;

  creating a signature of said sampled work based upon said plurality of segments; and

  storing said sampled work signature.

36.   The method of claim 35, wherein said method further includes the act of providing a plurality of reference signatures having a segment size and a hop size.

37.   The method of claim 36, wherein said method characterized in that said hop size of said sampled work signature is less than said hop size of said reference signatures.

38.   The method of claim 35, wherein said act of creating a signature of said sampled work comprises calculating segment feature vectors for each segment of said sampled work.

39.   The method of claim 36, wherein said act of creating a signature includes calculating a plurality of MFCCs for each said segment.

40.   The method of claim 35, wherein said act of creating a signature includes calculating a plurality of acoustical features from the group consisting of at least one of loudness, pitch, brightness, bandwidth, spectrum and MFCC coefficients for each said segment.

41.   The method of claim 35, wherein said sampled work signature comprises a plurality of segments and an identification portion.

42.   The method of claim 41, wherein said plurality of segments of said sampled work signature comprise a segment size of approximately 0.5 to 3 seconds.

43.   The method of claim 42, wherein said plurality of segments of said sampled work signature comprise a hop size of less than 50% of the segment size.

44.   The method of claim 42, wherein said plurality of segments of said sampled work signature comprise a hop size of approximately 0.1 seconds.
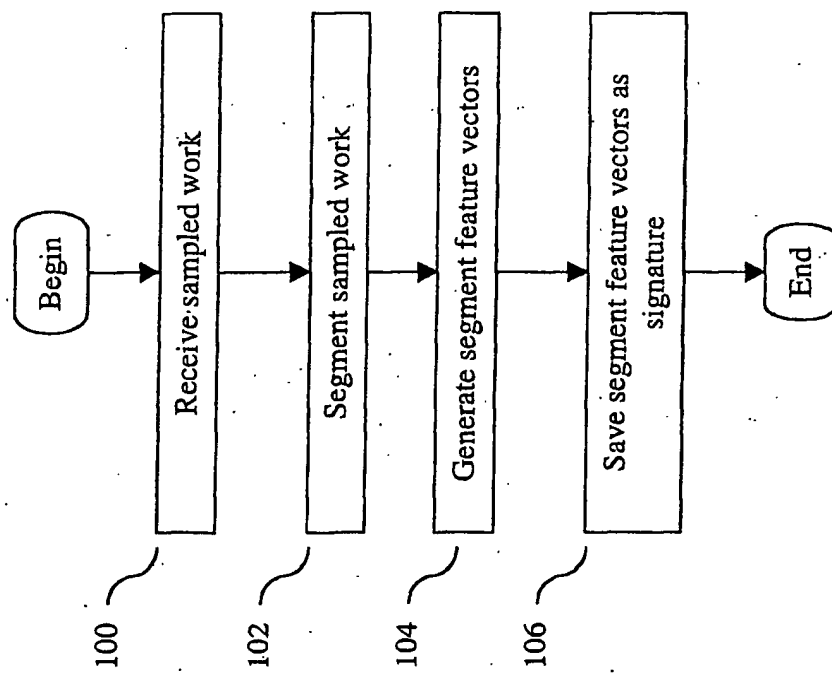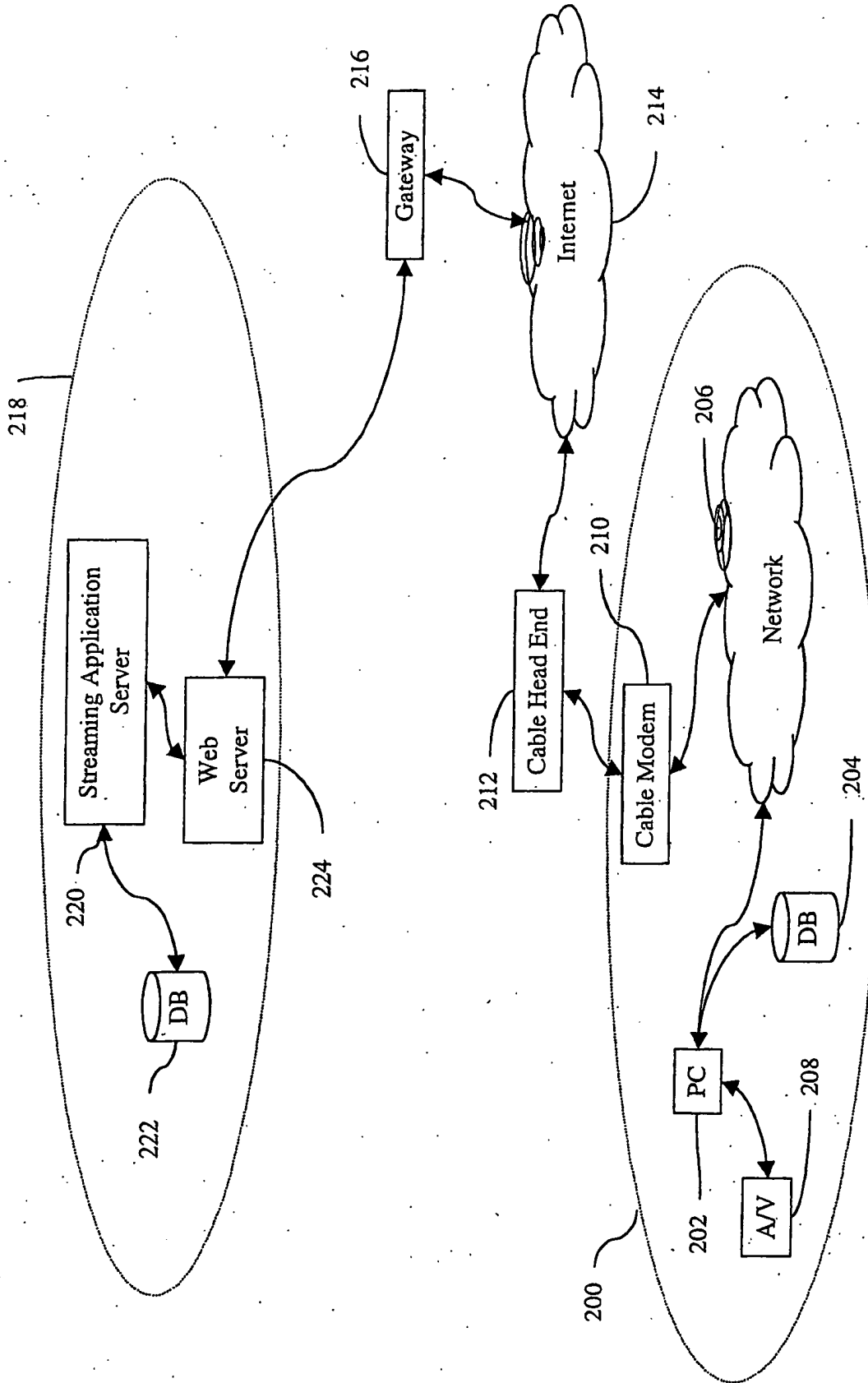
```
Begin

    │
    ▼

┌─────────────────────────┐
│  Receive sampled work   │  ⟶ 100
└─────────────────────────┘

    │
    ▼

┌─────────────────────────┐
│   Segment sampled work  │  ⟶ 102
└─────────────────────────┘

    │
    ▼

┌──────────────────────────────────┐
│ Generate segment feature vectors │  ⟶ 104
└──────────────────────────────────┘

    │
    ▼

┌──────────────────────────────────────┐
│ Save segment feature vectors as      │  ⟶ 106
│ signature                            │
└──────────────────────────────────────┘

    │
    ▼

End
```
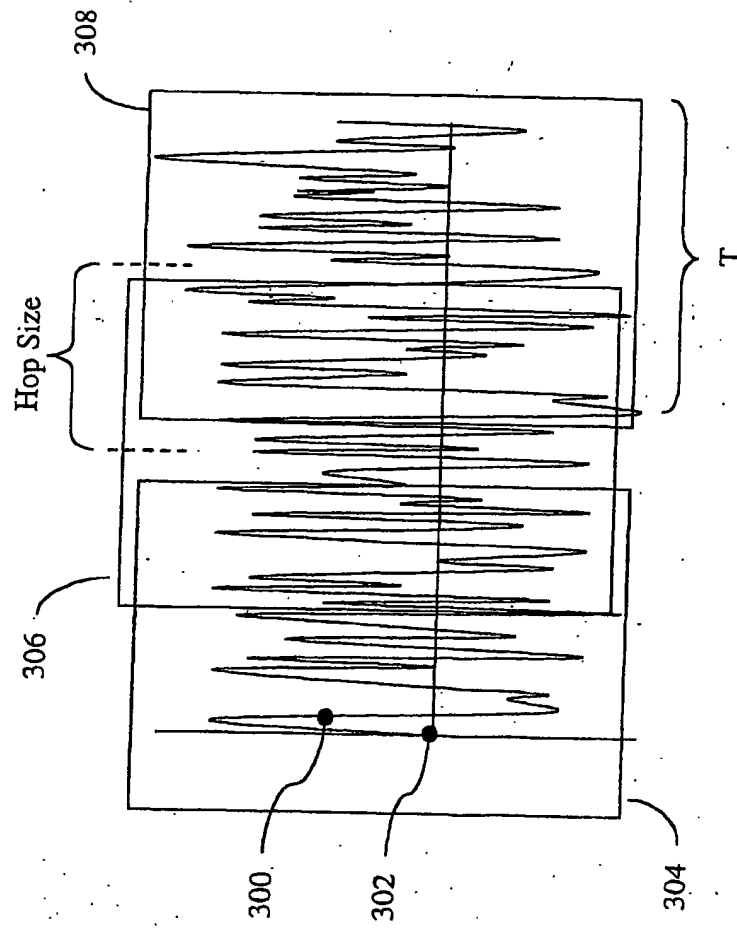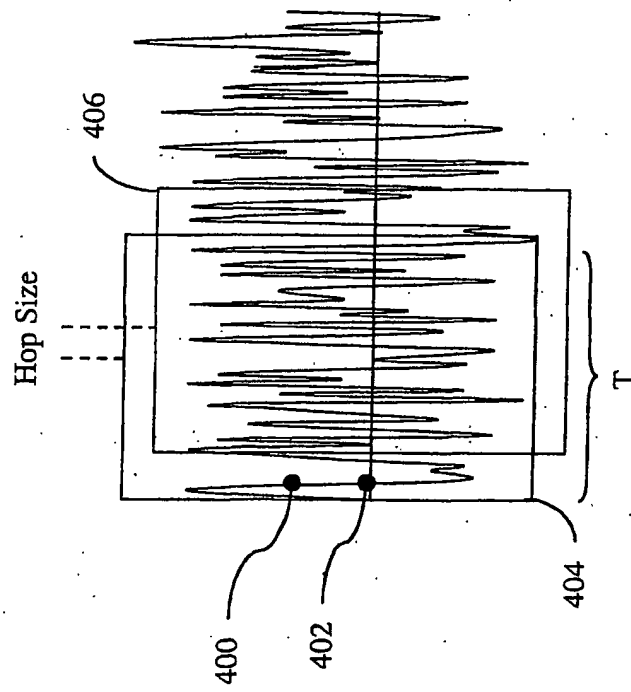
FIG. 1
Present Invention

FIG. 2
Present Invention

FIG. 3
Present Invention
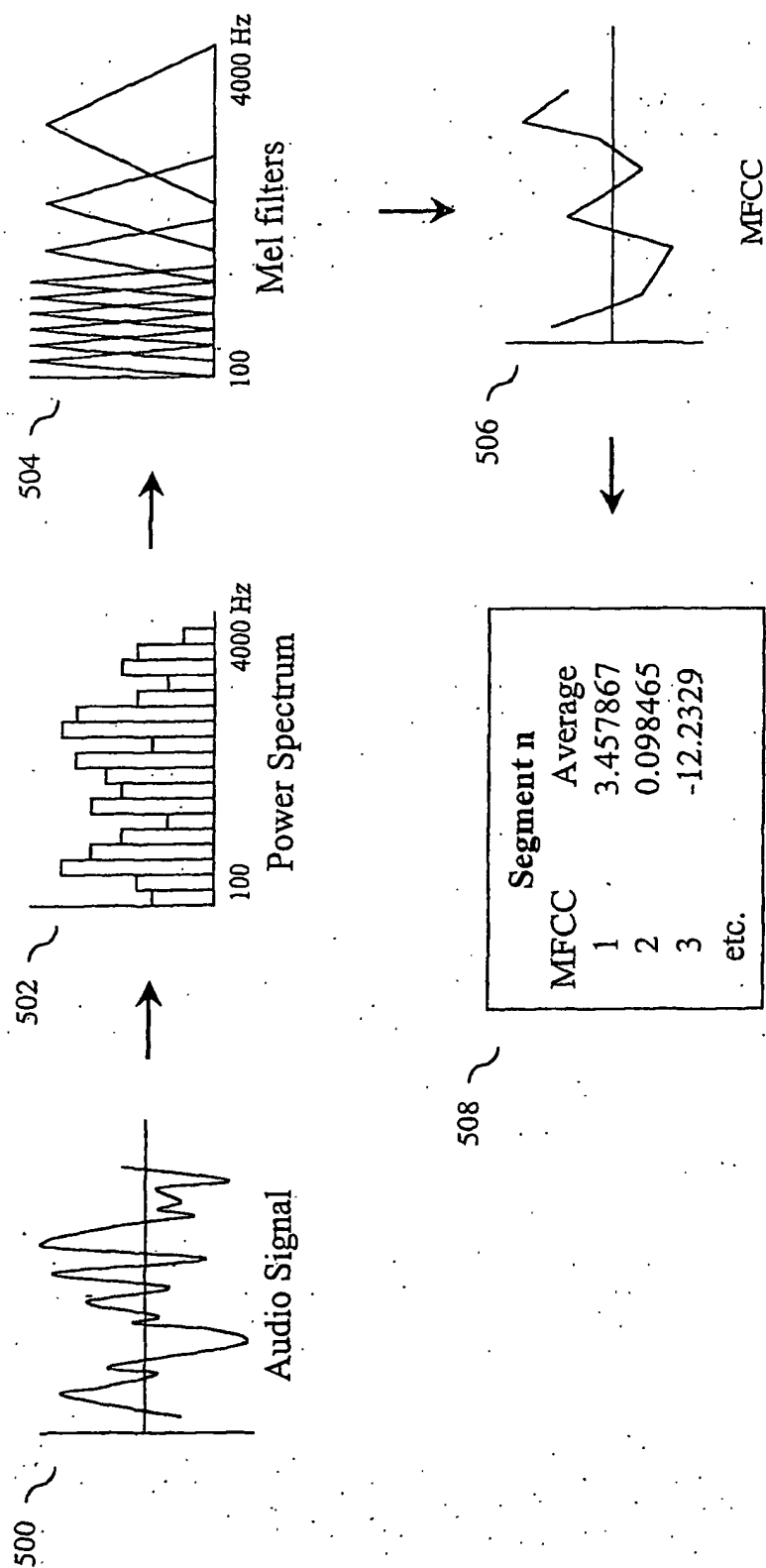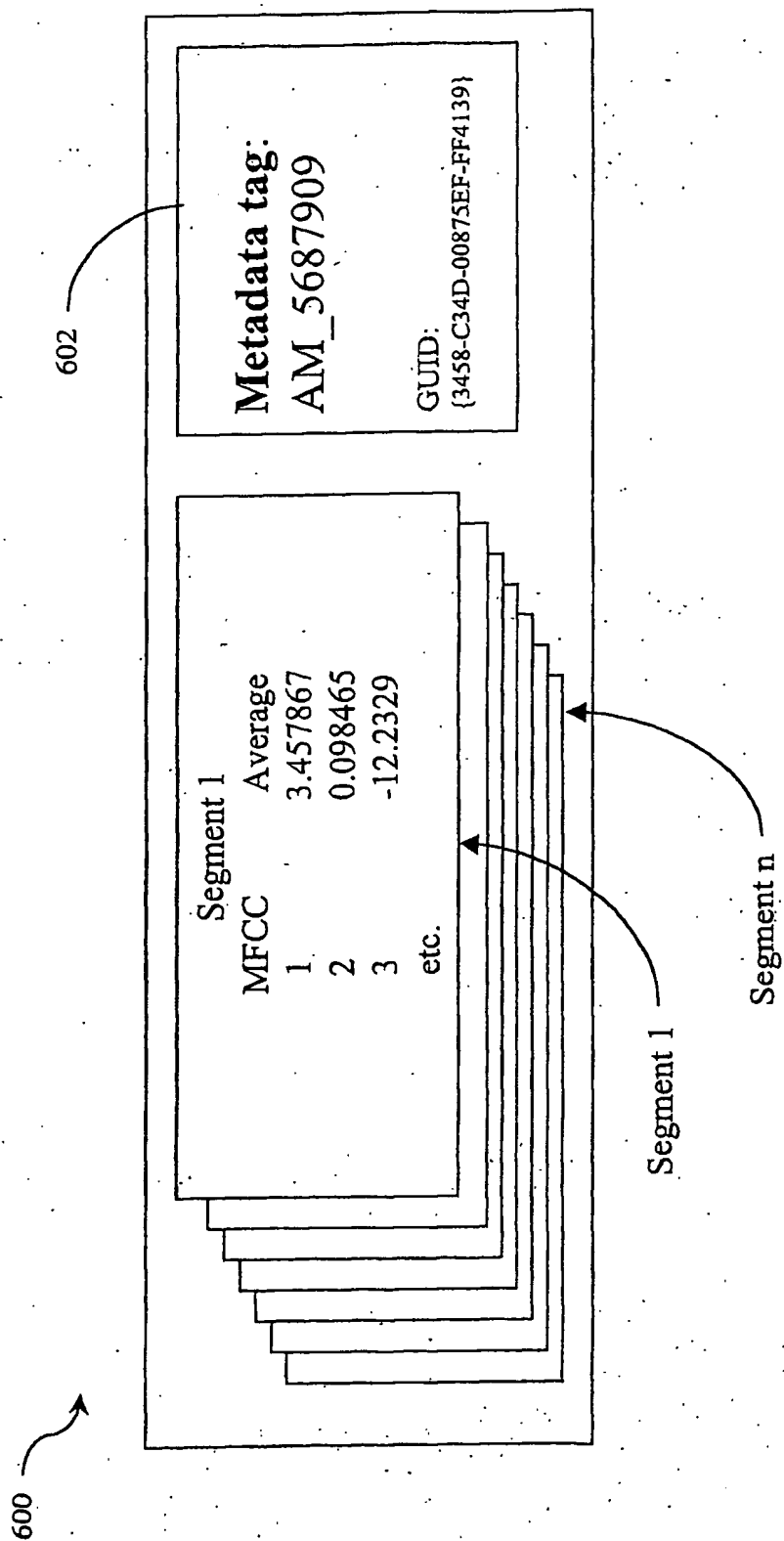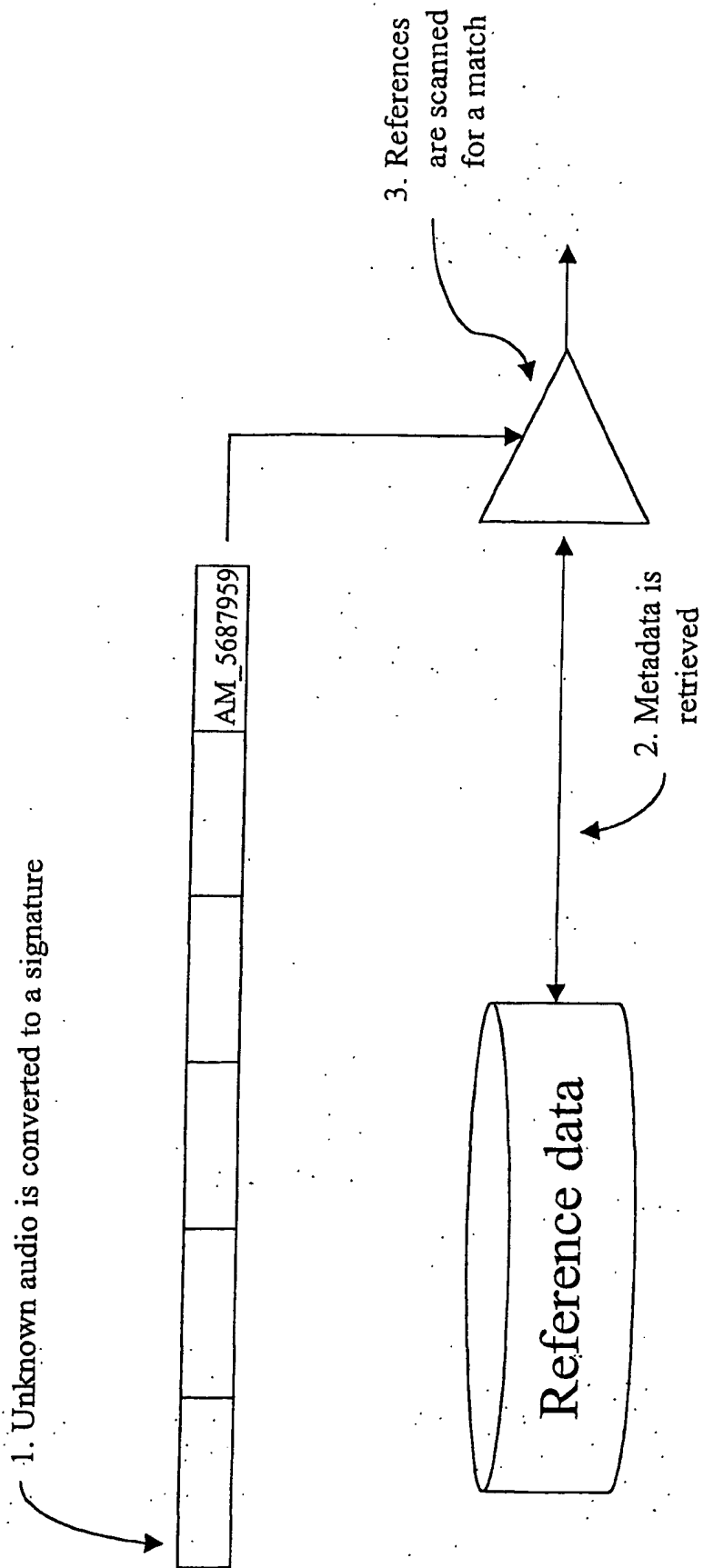
FIG. 4
Present Invention

FIG. 5
Present Invention

600

602

Metadata tag:
AM_5687909

GUID:
{3458-C34D-00875EF-FF4139}

Segment 1

MFCC   Average
1   3.457867
2   0.098465
3   -12.2329

etc.

Segment n

Segment 1

FIG. 6
Present Invention

1. Unknown audio is converted to a signature

AM_5687959

3. References
are scanned
for a match

2. Metadata is
retrieved

Reference data

FIG. 7
Present Invention